Food and Chemical Toxicology 65 (2014) 269–279

Contents lists available at ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

Development of a consumer product ingredient database for chemical exposure screening and prioritization

M.-R. Goldsmith [a,*], C.M. Grulke [a], R.D. Brooks [b], T.R. Transue [c], Y.M. Tan [a], A. Frame [a,e], P.P. Egeghy [a], R. Edwards [d], D.T. Chang [a], R. Tornero-Velez [a], K. Isaacs [a], A. Wang [a,e], J. Johnson [a], K. Holm [a], M. Reich [f], J. Mitchell [g], D.A. Vallero [a], L. Phillips [a], M. Phillips [a], J.F. Wambaugh [a], R.S. Judson [a], T.J. Buckley [a], C.C. Dary [a]

[a] United States Environmental Protection Agency (U.S. EPA), Office of Research and Development, RTP, NC 27711, United States
[b] Student Services Contractor at U.S. EPA, RTP, NC, United States
[c] Lockheed-Martin Information Technology, RTP, NC 27711, United States
[d] North Carolina State University, 2200 Hillsborough St., Raleigh, NC 27695, United States
[e] Oak Ridge Institute for Science and Education Fellow, United States
[f] University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, United States
[g] Biosystems and Agricultural Engineering, Michigan State University, E. Lansing, MI 48824, United States

ARTICLE INFO

Article history:
Received 16 October 2013
Accepted 19 December 2013
Available online 27 December 2013

Keywords:
Chemical exposure
Consumer products
Ingredients
Product formulation
Near field exposure
Exposure prioritization

ABSTRACT

Consumer products are a primary source of chemical exposures, yet little structured information is available on the chemical ingredients of these products and the concentrations at which ingredients are present. To address this data gap, we created a database of chemicals in consumer products using product Material Safety Data Sheets (MSDSs) publicly provided by a large retailer. The resulting database represents 1797 unique chemicals mapped to 8921 consumer products and a hierarchy of 353 consumer product "use categories" within a total of 15 top-level categories. We examine the utility of this database and discuss ways in which it will support (i) exposure screening and prioritization, (ii) generic or framework formulations for several indoor/consumer product exposure modeling initiatives, (iii) candidate chemical selection for monitoring near field exposure from proximal sources, and (iv) as activity tracers or ubiquitous exposure sources using "chemical space" map analyses. Chemicals present at high concentrations and across multiple consumer products and use categories that hold high exposure potential are identified. Our database is publicly available to serve regulators, retailers, manufacturers, and the public for predictive screening of chemicals in new and existing consumer products on the basis of exposure and risk.

Published by Elsevier Ltd.

1. Introduction

The production and use of chemicals is a hallmark of a modern and consumer oriented society. Within this modern society, however, there is also a growing awareness that there are thousands of chemicals to which humans are unavoidably exposed every day (Glegg and Richards, 2007; Sanderson et al., 2013; Weschler, 2009; Anastas et al., 2010). Of particular interest are chemicals from anthropogenic sources that can be potentially controlled. However, tracking and controlling these chemicals is a daunting task since it has been estimated that there are some 80,000 different chemicals in commerce (NTP, 2002). To evaluate the potential risk to human health associated with the use of these chemicals, there are two primary considerations: the chemical hazard and exposure. Because of the large number of chemicals and limited information about hazard or exposure, there is a need for screening methods of evaluation that have minimal data requirements and

Abbreviations: ACToR, Aggregated Computational Toxicology Resource; ADME, Absorption, Distribution, Metabolism, and Elimination; CAS-RN, chemical abstracts service registry number; CBI, confidential business information; CGI, common gateway interface; CHAD, Consolidated Human Activity Database; CPCPdb, Consumer Product Chemical Profile Database; CPSC, Consumer Product Safety Commission; CTCP, Clinical Toxicology of Commercial Products; DSSTox, Distributed Structure-Searchable Toxicity Database; EFH, Exposure Factors Handbook; HPDB, Household Product Database; HTML, HyperText Markup Language; IUPAC, International Union of Pure and Applied Chemistry; MSDS, Material Safety Data Sheet; NACE, Nomenclature des Activités Économiques dans la Communauté Européenne; NAICS, North American Industry Classification System; NEISS, National Electronic Injury Surveillance System; NLM, National Library of Medicine; OCR, optical character recognition; OSHA, Occupational Safety and Health Administration; Perl, Practical Extraction and Reporting Language; PHP, P Hypertext Preprocessor; REACH, Registration, Evaluation, Authorisation and Restriction of Chemical substances; SHEDS, Stochastic Human Exposure and Dose Simulation; SPIN, Substances in Products in the Nordic Countries; SOP, standard operating procedure; TSCA, Toxic Substances Control Act; URL, Uniform Resource Locator.

* Corresponding author. Tel.: +1 919 541 0497.
E-mail addresses: goldsmith.rocky@epa.gov (M.-R. Goldsmith), tan.cecilia@epa.gov (Y.M. Tan), isaacs.kristin@epa.gov (K. Isaacs).

0278-6915/$ - see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.fct.2013.12.029

can accommodate hundreds or thousands of chemicals at time (i.e., high throughput). Under a multi-agency initiative (NRC, 2007), considerable progress has already been made with respect to the development and implementation of high-throughput hazard testing. Progress on toxicity testing has in part highlighted the need for complementary high throughput evaluation of exposure.

For the purpose of high throughput exposure evaluation, there is strong justification based on both measurement (Morgan et al., 2005; Pellizzari et al., 1999; Wallace, 1987) and modeling (Lai et al., 2000; Nazaroff et al., 2012) studies to consider first and foremost those chemicals that are found in consumer products. These chemicals are brought into our homes where circumstances exist for relatively high levels of exposure depending on a chemical's prevalence in products and the frequency of product use along with chemical's degradation or removal rates (Hertwich, 2005). Indeed, many chemical ingredients that are reactive outdoors become relatively persistent indoors in light of physically different substrate (e.g. carpet twill with numerous sorption sites) and decreased biodegradation and photodegradation rates. More open systems, such as the ambient atmosphere, enhance a chemical's decay rate compared to the more closed indoor systems (Vallero, 2010). The indoor residential environment is where people spend a large majority of time, is where susceptible individuals (very young and elderly and those who are sick or disabled) tend to spend even more of their time (Klepeis et al., 2001), and is where most consumer products tend to be stored and used. It is recognized that depending on their purpose, exposure to chemicals in consumer products can result through either direct or indirect routes of contact (Schettler, 2006; Rudel et al., 2003; Weschler and Nazaroff, 2008). Historical examples demonstrating both the exposure and health risk potential of chemicals in consumer products that have led to their subsequent mitigation include lead in paint and chlorpyrifos in household pesticides.

Because consumer product use is an important determinant for human exposure to a broad range of chemicals, it follows that information about the chemicals and their concentrations within those products is also an important consideration. There are a few potential sources for such information. One is Material Safety Data Sheets (MSDSs) which are required under the Occupational Safety and Health Administration (OSHA) Hazard Communication Standard to include product ingredients with known toxicity. They are intended to inform workers and emergency response personnel of hazards and their safe management. All hazardous components in excess of 1% (0.1% for carcinogens) are required to be disclosed through product labeling and the MSDS. OSHA defines a hazardous chemical very broadly as one that could possibly cause any physical or health effect under expected conditions of use or reasonably anticipated conditions of misuse. OSHA does not require MSDSs to be provided to consumers. However, with greater public interest for this kind of information, many manufacturers and retailers are providing MSDS as a public service. For example, Walmart has made the MSDS inventory of their products available online (http://msds.walmartstores.com) for more than 10 years. A second source of such data is the Consumer Product Safety Commission (CPSC). The CPSC has been collecting data of this nature for internal use and regulatory purposes for the last four decades (Bracken and Weiss, 1977; Byer et al., 1976). A third source is the publication Clinical Toxicology of Commercial Products (CTCP), which was one of the first to aggregate consumer product ingredient compendiums dating back three decades; it included 14,000 products and over 900 generic product formulations (Gosselin et al., 1984). More recent efforts by the National Library of Medicine (NLM) provide similar data known as the Household Product Database (HPDB) in a web-accessible format indexed by consumer product category. The licensed form of this database is the Consumer Product Information Database (CPID). It includes the MSDS listing and chemical percent composition by weight when available and is soft-linked to multiple NLM informatics resources (http://whatsinproducts.com/). This database is unique in that it provides quantitative composition information which is critical for evaluating the exposure potential (Jayjock et al., 2008). A fourth effort is underway by the Chemical Trade Associations that represent both manufacturers and formulators to develop a Consumer Product Ingredient Communication Initiative to provide consumers with information about ingredients in products (Egeghy et al., 2011).

Accordingly, the current research is motivated both by need and opportunity. The large and growing number of chemicals that are used in consumer products result in considerable exposure potential to complex mixtures; therefore, there is a need for high throughput evaluation of exposure as a fundamental component of risk screening. At the same time, public information is increasingly available on which chemicals comprise a product. What is lacking is the methodology for capturing the available data in a form that supports high throughput exposure screening for rapid risk analysis (Goldsmith et al., 2012). In this paper, we rely on available data sources for collecting a structured set of product compositions that can be used to inform aggregate exposure and to generate generic product formulations used in regulatory exposure assessment models (i.e. EFAST or RIVM's ConsExpo). This structured set of product composition we have collated is the Consumer Product Chemical Profiles database (CPCPdb). The development of this database is well aligned with the National Research Council's report on exposure science (NRC, 2012) espousing the efficient development of data and the application of computational methods to derive new models that better predict exposure. Not only is CPCPdb a valuable tool for quantifying exposures, to aid in risk assessments, it is a scientifically credible means for evaluating exposure potential for a large number of chemicals leading to more effective prioritization of hazard characterization resources. Rather than focusing first on the chemicals that are known to have toxic effects and asking if there is a likelihood of exposure (i.e. hazard-centric risk assessment), the CPCPdb can be used to identify the chemicals with which the population has greatest contact so *in vitro* and *in vivo* toxicity testing can be used more judiciously. Exposure-based prioritization can be used to focus hazard identification and assessment efforts to a subset of relevant chemicals, streamlining the risk assessment process by reducing the dimensionality of possible (i.e. all chemicals) chemical exposures to the probable set of chemicals requiring our immediate attention.

2. Methods

The methodology for building the CPCPdb can be broken down into three major steps:

1. Building and curating a database for consumer product ingredients and percent compositions using available MSDSs.
2. Identifying and annotating product use categories for all products in the database.
3. Evaluating data quality.

Descriptions of each of these efforts are contained below and captured in Fig. 1.

2.1. Building and curating a database for consumer product ingredient and percent compositions using available MSDSs

2.1.1. Identification, retrieval, extraction

The first step necessary in the formation of the database was collecting consumer product MSDSs from a publicly available source. Walmart was chosen because (a) its consumer product portfolio is relatively diverse (e.g., personal care, automotive, arts & crafts, household pesticides, lawn/garden, cleaning, home maintenance, home improvement, office supplies, electronics); (b) Walmart's omnipresence in the consumer product marketplace is expected to provide a relatively accessible, high-market share product inventory with high consumer coverage; (c) its database of MSDS was available in Adobe PDF format; and (d) the documents were not only publically accessible, but also available without any visible

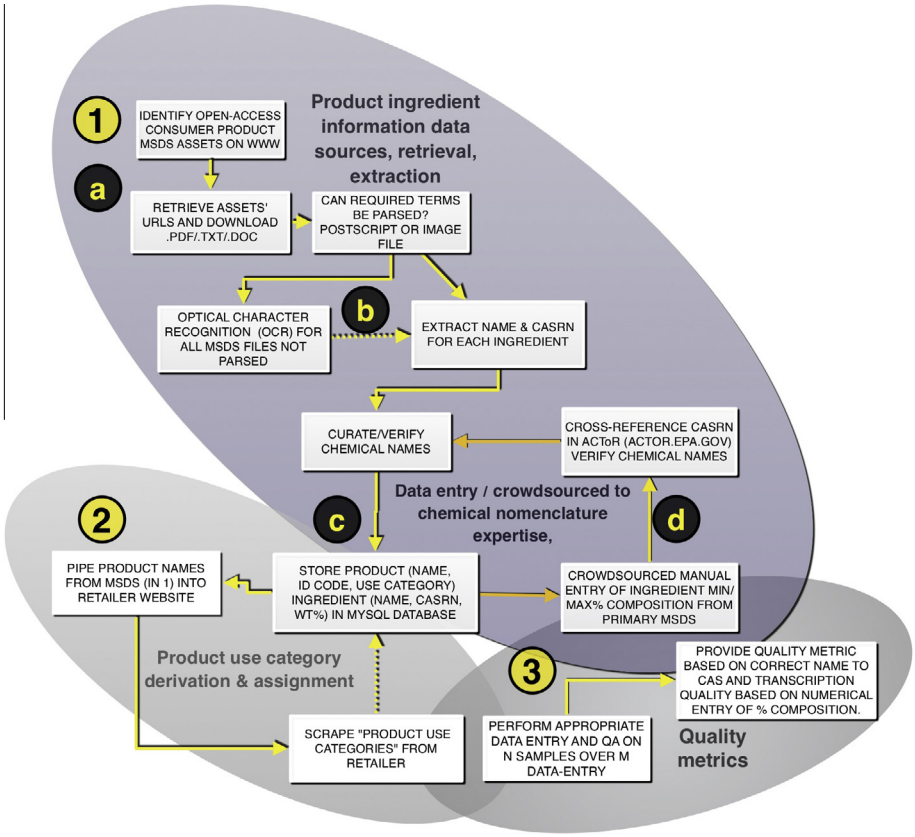

Fig. 1. Consumer Product Chemical Profile Data curation/annotation workflow: a total of ~10 K MSDS were identified, parsed, and manually annotated through a custom-built curation system (see Fig. 2). Details of the workflow steps 1–3 are explained in the methods.

restrictions. MSDS files were downloaded and entries were input into a MySQL (*MySQL: the world's most popular open source database*., MySQL AB, 1995) database documenting a sequential id, product name, file location, retrieval Uniform Resource Locator or URL, and document format.

2.1.2. Chemical ingredient data extraction and management

After download, programmatic tools were used to attempt to extract the ingredient information contained in each MSDS to the MySQL database. Roughly 70% of the MSDSs (in Adobe PDF) were embedded with text that was successfully parsed using custom scripts. The remaining PDF documents contained scanned images and were extracted using Tesseract-OCR [(a) Smith, 2007 and (b) https://code.google.com/p/tesseract-ocr/)] for optical character recognition (OCR) and subsequently annotated with appropriate regular expression filters. Entries for each identified ingredient in a given MSDS were crafted to document the MSDS in which it was found with chemical name and/or chemical registry number.

2.1.3. Chemical ingredient and quantitative composition data entry: crowd-sourced curation

Programmatic extraction of data from freeform or marginally formatted text is a science in its adolescence. To ensure that the CPCPdb had a level of quality that is commensurate with our need to support exposure assessment, a custom web-enabled interface was developed for manually curating every entry in our MySQL database (see Supporting Information 1A–C and Fig. 2 in Supplementary Data 2). The interface, developed in P Hypertext Preprocessor (PHP) and Practical Extraction and Report Language (Perl), provided curators/annotators with a means of opening a given MSDS, verifying the chemical names and associated registry numbers, and manually entering the "Min" and "Max" percent composition when available in the MSDS. A detailed standard operating procedure (SOP) was drafted and updated regularly to ensure that curators were following the same guidelines during mining and curation. An additional feature of this custom interface was its ability to identify products that contained identical chemical compositions or identical MSDSs (e.g., 32 oz versus 100 oz bottle of the same shampoo). The curator could then verify the level of congruence between the different MSDSs. If the MSDSs were identical, a "soft-link" between them was created. If chemical compositions in different MSDSs were similar, but not identical, the curator could copy curated information to reduce the amount of manual input needed. This feature reduced the overall workload of curation considerably since many products proved quite similar.

2.1.4. Chemical name verification

Chemical identifiers (e.g., common names, International Union of Pure and Applied Chemistry (IUPAC) names, registry numbers) selected either programmatically or via the manual curation interface were cross-referenced with chemicals listed in EPA's Aggregated Computational Toxicology Resource (ACToR) (Judson et al., 2012) (http://actor.epa.gov). Though both names and CAS-RNs were extracted from the MSDS, links were formed using CAS-RN to ACToR with names used only for verification and corroboration. Linking with ACToR allowed expansion of CPCPdb to include chemical structure so that secondary parameters could be derived that support a number of risk assessment objectives (i.e. Quantitative Structure–Activity Relationship, QSAR, based calculation of Absorption, Distribution, Metabolism, and Elimination (ADME) properties, physico-chemical properties, etc.). To ensure that the link was made correctly, the interface provided both the ACToR name and MSDS name that were then verified by either expert judgement or using appropriate secondary chemical name data sources, such as NLM's (http://pubchem.ncbi.nlm.nih.gov/) PubChem, the Royal Society of Chemistry's ChemSpider (www.chemspider.com), or the Distributed Structure-Searchable Toxicity Database (DSSTox) master list (http://www.epa.gov/ncct/dsstox/). For instance, if the MSDS listed "Ethyl alcohol" and the associated CAS-RN linked to an ACToR entry for "Ethanol", expert judgment was sufficient to identify that these are the same chemical and the entry verified that the chemicals were correctly linked. In contrast, an MSDS listing of "cetearyl alcohol", and an ACToR entry of "Alcohols, C16-18" resulted in a search of secondary sources. As curators became more familiar with the chemical composition of products and their related naming conventions, secondary sources became less necessary.

2.2. Identifying and annotating product categories for all products in the database

We chose to use the categorization scheme based on the retail product category given on the retailer's online shopping website (e.g., "beauty", "grocery", "home improvement") and a subcategory to which the product belongs (e.g., "shampoos", "pest control", "weed & fungus control"). This categorization scheme was already

somewhat integrated with the products data collected in this effort and provided a facile method for product organization. Further information of consumer product classification and its relevance to CPCPdb and exposure prioritization is available in Section 4.3.

A freely available data extraction program, ScreenScraper (http://www.screen-scraper.com/), was used to automatically retrieve product categories from the retailer's website (http://www.walmart.com). A program was coded to search the website for each product by name, and if a record was found, the associated categorical information was retrieved (see sections Supporting Information 1, Fig. 2 and Supporting Information 2 both found in Supplementary Data 2). Following the automated category retrieval, multiple manual checks were used to solve specific issues seen in the product classification. Misspellings and special characters in category names were corrected (i.e. things like "P$rsn%ol Care" or "personal car" were corrected to "Personal Care". Missing or clearly wrong categories were manually located via the retailers web interface. If the exact product was not found via the search, the expert curator made the decision whether the returned results were similar enough to be used as surrogates. If a similar product was not found, the category was eliminated. The software only returned the top record for a particular product search. Therefore, if a product was classified by the retailer in multiple categories, it was possible that very similar products could be assigned different categories by the automated search. To rectify this problem, curation experts evaluated every categorization to make sure that it was both consistent with assigned categories for similar products within our database and logical based on personal experience.

2.3. Evaluating data quality

Upon the completion of manual data curation and entry, we randomly selected approximately 130 MSDSs to check the accuracy of the process. Entries of each randomly selected product were compared to those in the MSDS. Omissions and transcriptional errors within the database for names, CAS-RN, and % composition were tabulated. Additionally, the stored categories for each of these products were evaluated based on both the results obtained by directly searching for the product on the retailer website and reasonable assumptions based on our experts' categorization knowledge.

3. Results

3.1. Data extraction

The results of our data extraction process are captured in Fig. 2. In total, 22 K MSDSs were retrieved from the retailer's public website. Initial extraction with natural language processing scripts yielded over 80 K unique MSDS/CAS-RN pairings with about 5% being retrieved from image based PDF. Further improvement of extraction scripts netted an additional 15 K putative MSDS/CAS-RN (with more than a third extracted from OCRed sources) giving a database with a total of 95 K unique MSDS/CAS-RN pairs. At this point, work on identifying products categories (Fig. 2I) for each MSDS was done in parallel with the manual curation effort (Fig. 2II) and then the results of each merged.

Manual curation identified 13 K duplicative entries among the 22 K MSDS entries (i.e. entries that have identical composition information as another MSDS entry) representing 5 K unique products. These duplicates are not unexpected and represent MSDSs for different packaging of the same brand of products (e.g., different sizes of the same formulation of toothpaste), MSDS for products that are identical but may contain different fragrances (e.g., perfumes, body washes), and MSDSs for competing brands (e.g., multiple rubber cement products with the same composition information). Taking into account duplicated entries, the number of unique MSDSs was reduced by 8 K (13 K removed and 5 K added back. This resulted in 27 K MSDS/CAS-RN pairings being removed as well. Of the roughly 68 K (80 K + 15 K – 27 K) remaining programmatically identified MSDS/CAS-RN pairings, 59 K of the pairings were confirmed to be correct by manual analysis (a specificity of roughly 87%). An additional 4.5 K MSDS/CAS-RN pairings that were missed by the automated procedure were also identified (indicating a sensitivity of 93% in automated CAS-RN extraction). In total, 13 K of the 14 K MSDSs were found to have extractable chemical information (name or CAS-RN). Of the 3.9 K unique chemical entries remaining after manual extraction, only 2.9 K mapped to a generic chemical in the ACToR database reducing the MSDS entries to 9.8 K.

Simultaneous to the extraction of chemical information from the MSDSs, product names from all 22 K MSDSs were used as queries in the website scraper script to identify their product categories. One of 433 product categories was identified for 20 K of

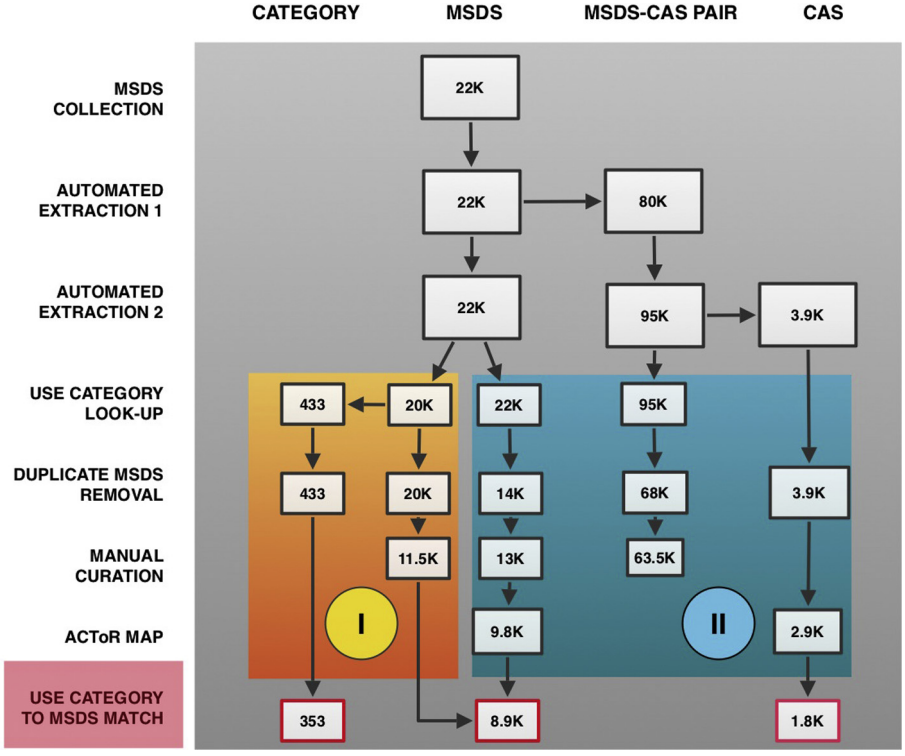

Fig. 2. Data extraction workflow.

these 22 K MSDSs. Following manual curation, we found only 11.5 K of the 20 K MSDSs were unique. When these 11.5 MSDS were matched with the 9.8 K MSDS that contained an ACToR generic chemical, only 8.9 K MSDS were found to map to a category and contained a valid chemical ingredient. From these MSDSs approximately 1.8 K unique substances were identified. Composition data (percent weight) was available for nearly 1.7 K of these chemicals. Using this methodology, we were able to categorize over 22 K product names (8.9 K unique products) containing nearly 1.8 K unique chemicals into 15 top level product categories, and 353 subcategories.

A complete flat-file (in Microsoft Excel .xlsx format) with "debranded" (anonymized) unique products (code), "product-use categories" (top and sub level), chemical ingredients (chemical identification number) and associated min/max % compositions is available in the Supplementary Data 1. This dataset does not include chemical names extracted that could not be linked to their CAS-RN nor entries for which quantitative composition information is unavailable. These additional data will be contained in future releases of CPCPdb on ACToR (http://actor.epa.gov).

3.2. Quality assurance (QA)

As CAS registry numbers were considered the most definitive identifier for a chemical during ingredient extraction, it was one of the key pieces of information included in our database. In the 130 product subset for quality assurance, only 7 of the products had a case where there was a listed CAS-RN in the MSDS that was not identified in our database. Only 1 entry had improperly transcribed CAS-RNs. In total, 97% (587/608) of the CAS-RNs in the 130 product sample were properly annotated. Secondary to CAS-RN extraction was the collection of chemical names listed in the MSDS. CAS-RNs were typically found in the tabulated ingredient section in the MSDS; however, names were found in nearly all sections of the MSDS making omission more likely. Upon detailed review of the 130 selected MSDSs, 88% (114/130) were found to have no missed chemical names. For chemical names, 824 of 914 names listed on the 130 MSDSs were captured. While the collection of names was more error prone than CAS-RN identification, the specificity of ingredient identification was errorless (i.e. no names were listed that did not occur in the MSDS, or conversely, all names listed were in fact in the MSDS). This high specificity is key since improperly labeling a product as containing a chemical is more problematic than missing an ingredient that does occur (Note that ingredient listings from MSDSs already have a relatively low sensitivity due to the limitations of legal requirements of MSDS reporting).

During QA, we found that more than 80% of compositions identified within the MSDS were annotated in the database. There were only a small number of cases (1 MSDS with 5 ingredients) where composition data stored in the database was incorrect. These errors were due to transcriptional errors in the proper linking of ingredient with quantitative composition; ingredient and max concentration values were mismatched due to ordering difference in the QA interface and the MSDS.

QA was also completed on the categorization of the products listed within the CPCPdb. Based only on comparing to the category of the first record returned on the company website, 87.5% of the product categories matched at the top-level (15 categories), and 73% matched at the most detailed level (353 categories) of the classification tree. These differences were mainly caused by changes made by the retailer's website during the lag time between the original annotation (mid 2012) and the QA (Early 2013). Some categories that we had in the database no longer appear on the retailer's site. Some products were no longer found on the website or were found in more than one category.

The omission of an MSDS/ingredient pairing in our data is of concern as it could lead to underestimating exposure to a particular chemical. However, when performing exposure assessments, it is more common to examine a generic formulation for a product category rather than the composition of a single product. When considering the errors of omission identified in the ingredient database, we consider it unlikely that the same ingredient would be mistakenly omitted from all products contained within a single product category. For example, a hand soap (product #1528 in the Supplementary Data 1) falls under the "Health: Hand Soap and Sanitizers" retail product code category. This category also contains an additional 105 products spanning 100 unique chemicals. Each of these 100 chemicals are present in at least 2 (although mostly $\gg$2) products within the category (see Supporting Information 1D in Supplementary Data file 2), or more generally speaking, these chemicals constitute the *framework* or *generic formulation* for that product use category. If a chemical or percent composition for a particular product was not captured in the curation process, the chemical would still be captured in other similar products within the same *generic formulation* (see Supporting Information 1E for framework/generic formulation analysis in Supplementary Data 2).

3.3. Ingredient diversity

We observed wide variability in chemical ingredients across products. A more comprehensive evaluation of chemical prevalence can be gained through a network diagram analysis (Fig. 3) with the large nodes representing top-level product categories and the small nodes representing unique chemical species. Edges are drawn if a chemical is present within a product in that product category. In looking at Fig. 3 (generated using Cytoscape: Shannon et al., 2003). It is easy to see three different classes: chemicals that are only present in a single product class (in green), those that are present in a small number of classes (in orange), and those that exist in a large number of product classes (in red).

The foremost class (containing compounds that are only present in a single product category) contains chemicals that may be used as *sourcetracers* for a specific product category. Take for example lithium hydroxide, which is only found in beauty products–particularly hair conditioners and moisturizers. If this chemical is measured within the home, blood or urine, the CPDBdb could be used to identify its likely consumer product source. Alternatively, one could use CPDBdb in combination with use and physical/chemical characteristics to predict exposure. Other *sourcetracer* chemical examples include 3,6,9-trioxaundecamethylene bis(2-ethylhexanoate), which is only present in"Home Improvement: Sealants, Fillers and Adhesives"; 2,5-thiophenediylbis (5-tert-butyl-1,3-benzoxazole), which is only in "Toys: Drawing and Coloring";Phosphordithioic acid, which can be found in "Automotive and tires"; and ethofenprox, which is a chemical in "Pets: Flea and Tick Control". Furthermore, by knowing the degradation properties of sentinel chemicals, it may be possible to determine the likely time a product was used. This type of forensic examination may allow us to harness chemical markers to tell us about product usage in the home without relying on somewhat error-prone surveying mechanisms to monitor product application patterns. In addition, with the large number of ingredients identified to be potential tracers, scientists can choose the chemicals with the optimal properties for the detection media they prefer.

The second class, colored **orange**, (containing compounds found in **multiple** product categories) contains chemicals that have a lower value as a tracer but may be of greater concern with respect to exposure since their presence in more products could lead to higher aggregate exposure.

The final class, colored **red,** (containing compounds which are found in **most** product categories) contains chemicals with no use

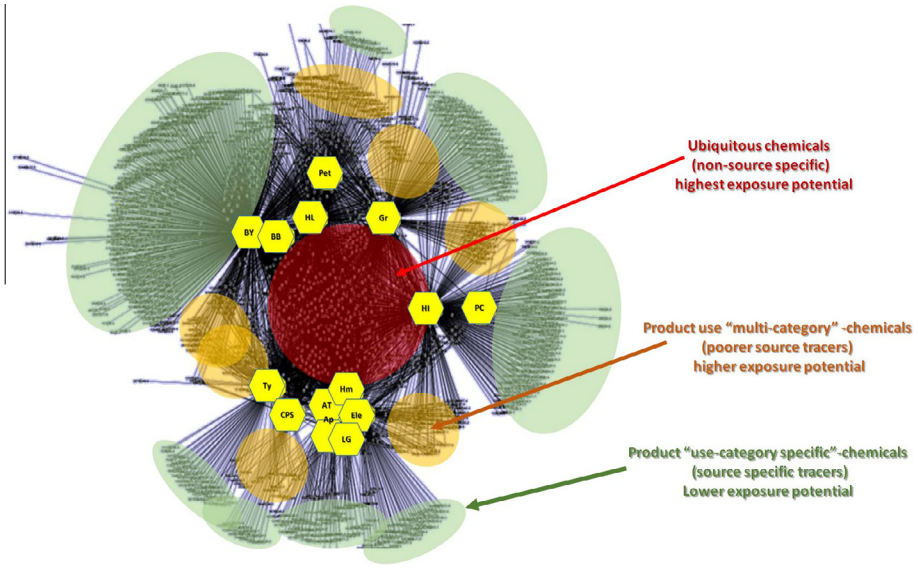

Fig. 3. The chemical space of consumer products: network diagram of 15 top level *product use categories* nodes (yellow) connected to ~1.6 K chemicals and summary table of the CPCPdb product and chemical landscape. The graph is highlighted with various individual chemical entities (short text strings in black) with regions of interest colored according to use as tracers, and their exposure potential. Product category nodes (yellow) with similar "ingredient space" (or chemical profiles/signatures/formulations) cluster closer together. (Legend: BY = beauty, BB = baby, HL = health, Pet = Pet care, Gr = Grocery, HI = home improvement, PC = photo center, Hm = home, Ele = electronics, LG = gardening, Ap = Apparel, AT = automotive and tires, CPS = crafts and party supplies, Ty = Toy, Sports and outdoor hidden under AT, Ele, LG) A lower product category list to CASRN exists as a downloadable Cytoscape file for network exploration in Supplemental Information 4 found in Supplementary Data 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

as tracers, but have the greatest exposure potential. Many of these ubiquitous chemicals have been well studied; out of a total of $n = 34{,}345$ structure-product pairs the top 10 chemicals in terms of frequency are water (n = 2,005, average weight % = 58.3), ethanol ($n = 1130$, average weight % = 35.4), glycerol ($n = 1043$, average weight % = 8.2), titanium dioxide ($n = 777$, average weight % = 7.6), 1,2-propanediol ($n = 600$, average weight % = 6.9), methyl 4-hydroxybenzoate (methyl parabens, $n = 516$, average weight % = 0.4), propane ($n = 470$, average weight % = 10.8), propyl 4-hydroxybenzoate (propyl parabens, $n = 456$, average weight % = 0.4), isobutane ($n = 398$, average weight % = 2.3), isopropyl alcohol ($n = 390$, average weight % = 16.1). The prevalence of chemical use in consumer products can be used as a basis for screening chemical exposure or if combined with additional information (e.g. prevalence and type of use, fugacity, environmental fate and transport), provides a basis for a rarefied exposure assessment.

The similarity/diversity of ingredient profiles in different product classes is also evident in Fig. 3. Products with similar ingredients cluster closer together (for instance "beauty", "health" and "baby", "grocery" are closer together than a satellite product nodes of "apparel" and "photo center". The remaining classes then cluster together. This could be important in proper selection of a categorization scheme that would be usable in exposure prioritization efforts. The greater number of categories, the more difficult a exposure prioritization model will be to parameterize. However, if ingredient profiles are similar for multiple product classes, it may be possible to consolidate those classes under a single modellable category. Using ingredient knowledge (along with usage and exposure route similarities) to inform the categorization process could lead to improved near field exposure models that can accurately evaluate exposure potential with minimal parameterization burden.

3.4. Biomarker-CPCPdb overlap

Based on the results from examining our network diagram (Fig. 3), the CPCPdb may help to identify tracer chemicals and biomarkers of exposure. By examining unique chemical signatures or footprints resulting from product use or activity patterns in conjunction with *in silico* derived chemical property filters, estimates can be made as to the persistence and partitioning of substances within the environment and/or the body (Kelly et al., 2007; Sarver et al., 1997; Berellini et al., 2009; Weschler and Nazaroff, 2010). For example, a chemical with a (relatively) long half-life and low vapor pressure can be predicted to occur within dust long after its use. Such analyses could be used to identify candidate source signatures for consumer products in environmental media such as dust or to identify chemical candidates for surveillance programs such as the NHANES biomonitoring program (Shin et al., 2013). NHANES chemicals (http://www.cdc.gov/nchs/nhanes.htm) are roughly grouped into chemical classes (e.g. environmental phenols, herbicides, perfluorinated compounds); among those chemical classes likely to be in consumer products (listed in Table 1), the biomarkers monitored by NHANES imply exposure to at least 210 different chemicals.

As is shown in Fig. 4 and Table 1, of the 210 NHANES chemicals, only 47 appear in the CPCPdb, however we capture a chemical ingredient in over 70% of the unique chemical classes (14 of 20 classes). Parabens, Phthalates, and Volatile Organic Compounds (VOCs) are in the largest number of product categories. Beyond these frequently used chemicals, Fig. 4 shows that most chemicals are in products with only a handful of uses, indicating that the presence of biomarkers for exposure to these chemicals may correlate with specific consumer product-related activities. Bisphenol A, flame retardants (Polybrominated and Diphenyl Ethers) and perfluorinated compounds are not found, though they are known to be either in packaging or used in the manufacture of many products in the home.

4. Discussion

The development of CPCPdb is notable with respect to both the creation of the database as well as defining the methodology. Information on chemicals in consumer products is an important consideration for high-throughput methods to infer near field

Table 1
The occurrence of chemicals within the CPCPdb of various NHANES chemical classes that might be in the home.

			Product categories			
			Lower categories		Upper categories	
Class	Number of compounds	Number in CPCPdb	Number	Most common	Number	Most common
Carbamates	6	1	2	Flea and Tick	2	Grocery
DEET	1	1	3	Cleaners	2	Grocery
Dithiocarbamate pesticides	8	0				
Environmental phenols	4	2	21	All Fragrances	2	Beauty
Herbicides	7	3	4	Fertilizer and Soil	3	Home Improvement
Organochlorine pesticides	21	1	3	All-Purpose Cleaners	2	Grocery
Organophosphate pesticides	2	1	2	Pest Control	2	Grocery
Organophosphorus insecticides	39	7	2	Pest Control	2	Grocery
Other pesticides	7	1	9	Pest Control	4	Pets
PAHs	12	3	5	Air Fresheners and Deodorizer	3	Grocery
Parabens	4	4	75	Accessories	12	Beauty
Perfluorinated compounds	12	0				
Phthalates	10	7	23	Playsets	9	Beauty
Phytoestrogens	11	0				
Polybrominated Biphenyl	1	0				
Polybrominated diphenyl ethers	10	0				
Pyrethroid pesticides	4	4	9	Pest Control	5	Pets
Sulfonyl urea herbicides	17	0				
Tobacco	1	0				
VOCs	33	12	54	Sealants, Fillers and Adhesives	11	Home Improvement
Total	210	47				

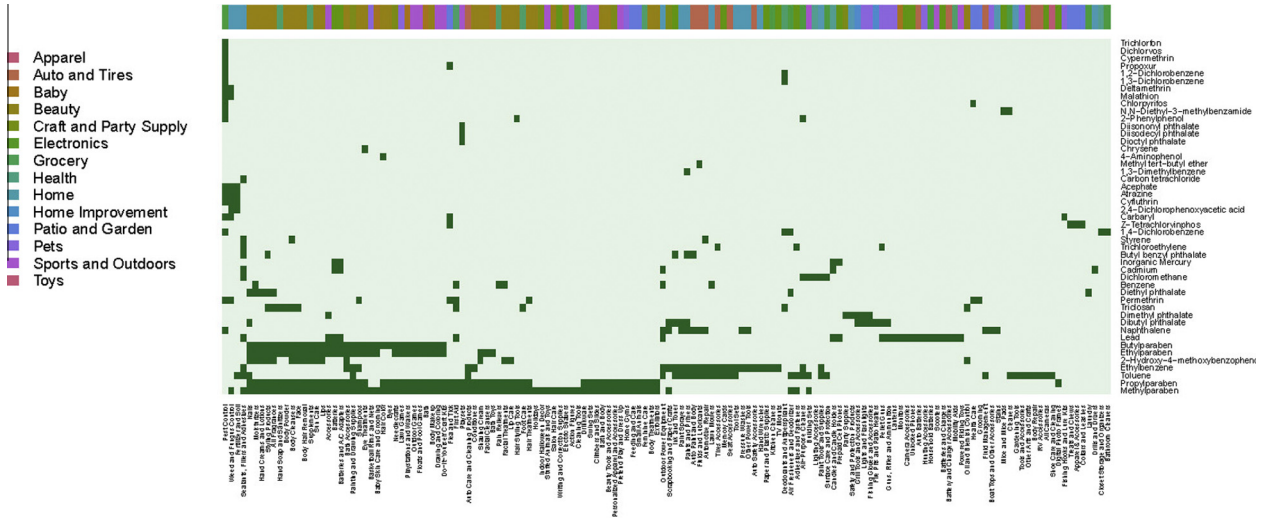

Fig. 4. Product categories for NHANES Chemicals that are in the CPCPdb. Dark color indicates the presence of these chemicals (rows) in a "specific" product (lower) category. The ribbon at the top indicates the corresponding upper product category (14 of 15 categories). Chemicals and products were clustered using Euclidean distance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

exposure for most chemicals (Little et al., 2012; Nazaroff et al., 2012; Shin et al., 2012; Wenger et al., 2012). Recent evaluation of high-throughput exposure methods demonstrates that near field use is highly predictive of exposure (Wambaugh et al., 2013). CPCPdb provides scientists and regulators with the means to better estimate potential near field exposures. The information in the CPCPdb bridges a key data gap that limits the characterization of real-world exposures. These (and many other diverse types of data) are necessary to quickly and accurately assess exposure potential to complement toxicity screening in support of rapid risk assessment.

The likelihood of a person being exposed to a particular chemical is dependent on multiple factors, but that chemical being an ingredient in a consumer product is certainly an important variable. CPCPdb is intended to document the diversity of chemical ingredients within the multitude of products commonly purchased by consumers. Other data sources such as the NLM's Household Product database include on the order of 3000 chemicals, however many of the products listed there are now discontinued (nearly 30%). It is vital to keep the chemical space covered by consumer product ingredients current as products may have been reformulated or discontinued due to health concerns of certain chemicals, product efficacy optimization, environmental impact, or better/cheaper alternatives being discovered. If consumer product ingredient database used during an exposure estimation still includes such products, we could greatly overestimate the public exposure to chemicals in those products and identify "risks" that in reality have already been mitigated. It is therefore vital that the database be updated frequently with new products, but equally important that outdated data be removed from the database. The vision for

CPCPdb is an evergreen database that will be regularly updated and pruned to ensure its utility to exposure scientists.

Currently, our database contains roughly 1800 chemicals, but it is important to note that CPCPdb is based on MSDS forms due to their availability and parsability. Naturally, there are limitations in the data MSDS provide that could lead to underestimating exposure and risk. There are many reasons the CPCPdb may not document a chemical that is present in near field products: (1) some chemicals are trade secrets, (2) consumer product may contain chemicals that are not ingredients, (3) our database was built using the inventory of a single retailer, and (4) MSDS are typically not available for articles (an important near field exposure source). These concerns along with difficulties in consumer product categorization and additional uses for CPCPdb are expounded upon *infra*.

4.1. MSDS omissions

In an effort to safeguard intellectual property and protect trade secrets, the formulation of a product and the choice of certain chemical ingredients, are typically classified under confidential business information (CBI). Not all ingredients are required to be reported in an MSDS. Obtaining information on fragrance and phthalate ingredients is known to be particularly problematic (Egeghy et al., 2011). For instance, there are well over 3060 unique materials used in fragrances, chemicals that have received considerable scrutiny due to their abundance and diversity (several thousands are used in various combinations in several thousands of products), associations with asthma and multiple chemical sensitivity, yet it is completely acceptable to include them in consumer products without listing a CAS-RN or a name (i.e., simply "trade secret"). We have not stored these components as it is impossible to identify them or even infer what they might be (e.g., one product contained twelve ingredients, each listed as "trade secret" with a percent composition listed as "1–100%").

While it is true that many manufacturers are now reporting more of their product ingredients due to ever increasing public pressure for transparency, there is still a wide variability in reporting behavior and possibility of omissions must be considered in usage of the data. One potential way to counteract omission is to use the large sample size within CPCPdb which contains many brands of products for any given consumer product class. While chemical information on certain individual products may be lacking, the aggregated data across a product class should effectively capture the majority of possible ingredients. This is verified by the analysis in 3.2 on how omissions in our extraction process can be overcome using the other data available in CPCPdb. The effectiveness of using composite ingredient data for a specific category supports an informatics-driven approach to rapidly summarize product-use categories into generic or framework formulations or chemical profiles; a requirement to reduce the chemical space independent of product branding, and a data requirement for many consumer product exposure models today that require generic formulation to reduce dimensionality of exposure models. Such categorical usage of the data provides more consistent results and overcomes the limitations caused by omissions regardless of their source.

4.2. Consumer product non-ingredients

In addition to ingredients that are covered under trade secret protection, there are also chemicals that can be detected in consumer products that are not knowingly added such as accidental contaminants, products of degradation, by-products of manufacturing or chemical synthesis and residues from packaging or handling equipment. Since these chemicals are not intentionally added they are not documented in MSDS and are not included in our database, although some could potentially be inferred from product packaging material labeling or an RIC (resin identification code). For instance, if a product is purchased in a bottle that is either PVC or polycarbonate or other resin, chemicals used to produce the plastic resin may inadvertently become part of the product formulation. Important examples include phthalates, which are used as plasticizers in a variety of otherwise brittle polymers, and bisphenol A (BPA), which is the monomer form of some polycarbonate polymers. It is highly unlikely that the manufacturer would list these chemicals, as they are a result of the production and packaging process and are not actual "ingredients." Despite not being explicitly listed as present, phthalates are commonly detected in bottled water because of the material used in the bottle or other packaging (Al-Saleh et al., 2011).

The Silent Spring Institute recently released a study that started with a list of 66 chemicals of interest because of their putative adverse health outcomes (Dodson et al., 2012). The study detected 55 of these chemicals in 213 different consumer products; however, several of these chemicals are not true ingredients, but are rather by-products, accidental contaminants, or packaging residues. However, the manufacturer of a consumer product does not necessarily also manufacture the packaging and therefore may be unaware of additional ingredients potentially introduced by packaging. Only by collecting both sets of information (packaging and ingredient data) can the full diversity of the exposure due to consumer products be evaluated.

4.3. Coverage of our MSDS collection

The current release of CPCPdb only contains "consumable" products (e.g. liquids, solids, aerosols, powders) for a single retailer. There are many more retailers with different market focuses that deal in specific types of consumer products (e.g. home improvement, auto parts, art/craft). While the retailer we chose covers a broad range of products, its lack of focus on specific areas may lead to underestimation of the diversity of chemicals to which practitioners of specific hobbies may be exposed. It is our desire to expand our MSDS collection to include those made available by other retailers.

In addition, by using MSDS as our information source, an important class of near field products is missing in our database—articles (e.g. electronics, furniture, building materials). Identifying and characterizing the chemicals found in articles and their corresponding exposure potential (via emission into indoor air or via human contact) is an ongoing area of research. (Little et al., 2012; Nazaroff et al., 2012; Shin et al., 2012; Wenger et al., 2012; Wilke et al., 2004).

4.4. Consumer product categorization

Consumer product categorization is necessary for developing generic product formulations and has proven to be valuable in analysis of the data within the CPCPdb. However, there is still much work to be done in finding an appropriate categorization scheme for consumer products in the context of exposure assessment.

More classical uses for product classifications are informing manufacturers of the strategies needed to compete for market share and providing the basis for the society to monitor the health of the economy through observation of trends within specific sectors (Meyer & Slick, 2001). It has also been shown that a system for product classification is valuable in assessing exposure to chemicals in those products during their manufacture, processing, and use (OECD, 2012). As a result, many categorization schemes are available for classifying consumer products (Table 2). These schemes generally fall into three categories: "industrial function",

Table 2
Common schemes for the classification of consumer products, functional categories, economic sectors, and related activities and locations.

Abbrev.	Name	Sponsor	Availability
Product categories			
HPDB	Household Products Database	U.S. National Library of Medicine	http://householdproducts.nlm.nih.gov/index.htm
C Series	U.S. – Canada Harmonized Consumer and Commercial Product Code	U.S. EPA, Health Canada, Environment Canada	http://www.epa.gov/cdr/tools/InstructionsManual.013112.pdf
EFH	Exposure Factors Handbook Consumer Products	U.S. Environmental Protection Agency	http://www.epa.gov/ncea/efh/pdfs/efh-chapter17.pdf
NEISS	National Electronic Injury Surveillance System	U.S. Consumer Product Safety Commission	http://www.cpsc.gov//PageFiles/106513/completemanual.pdf
Simmons	Simmons National Consumer Study	Experian Marketing Services	http://www.experian.com/simmons-research/consumer-study-details.html
REACH PC/AC	REACH Use Descriptor System Chem-ical Product and Article Categories	European Chemicals Agency	http://echa.europa.eu/documents/10162/13632/information_requirements_r12_en.pdf
UCN	Use Code Nordic	Nordic Council of Ministers, Chemical group	http://90.184.2.100/DotNetNuke/LinkClick.aspx?link=DNNPortal-Download%2fFunktionskoder-eng+htm.htm&tabid=58&mid=448
SPIN/ UC62	Substances in Products in the Nordic Countries	Nordic Council of Ministers, Chemical group	http://90.184.2.100/DotNetNuke/LinkClick.aspx?link=UC62_ExplanatoryTxt.doc&tabid=58&mid=448
Function/process categories			
U Series	U.S. – Canada Harmonized Industrial Function Categories	U.S. EPA, Health Canada, and Environment Canada	http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono%282012%295&doclanguage=en
PROC	REACH Use Descriptor System Process Category	European Chemicals Agency	http://echa.europa.eu/documents/10162/13632/information_requirements_r12_en.pdf
Appendix R.12-6	REACH Use Descriptor System Functional Category	European Chemicals Agency	http://echa.europa.eu/documents/10162/13632/information_requirements_r12_en.pdf
Economic activities			
NAICS	North American Industry Classification System	U.S. Census Bureau, Statistics Canada, and Mexico's Instituto Nacional de Estadistica y Geografia	http://www.census.gov/eos/www/naics/2012NAICS/2012_Definition_File.pdf
IS	Industrial Sector	U.S. Environmental Protection Agency	http://www.epa.gov/cdr/tools/ReplacingNAICSwithIS.pdf
NACE	Nomenclature des Activités Économ-iques dans la Communauté Européenne	European Commission Directorate for Economic and Financial Affairs	http://ec.europa.eu/competition/mergers/cases/index/nace_all.html
SU	REACH Use Descriptor System Sector of Use	European Chemicals Agency	http://echa.europa.eu/documents/10162/13632/information_requirements_r12_en.pdf
Activity and location			
CHAD	Consolidated Human Activity Database	U.S. Environmental Protection Agency	http://www.epa.gov/heasd/chad.html
ATUS	American Time Use Survey	Department of Labor, Bureau of Labor and Statistics	http://www.bls.gov/tus/

"consumer" and "commercial product". The U.S. Environmental Protection Agency (EPA), Health Canada, and Environment Canada recognize have developed harmonized codes to support communication and application for exposure characterization. Similarly, a crosswalk with the function and product codes in the Use Descriptor System developed by the European Chemicals Agency to support the implementation of Registration, Evaluation, Authorization and Restriction of Chemical substances (REACH) Regulation was developed under the auspices of the Organization for Economic Development (OECD, 2012). As these classification schemes are designed both to reduce burden on industry in chemical inventory reporting and to minimize the number of setting/process/use combinations that comprise the exposure scenarios used for evaluating chemicals, they broadly consolidate products into more general categories. The more detailed and hierarchical categorization schemes either tend to describe economic activities rather than products (e.g., North American Industry Classification System, NAICS, Nomenclature des Activités Économiques dans la Communauté Européenne or NACE), or share little consistency across applications (e.g., HPDB, National Electronic Injury Surveillance System (NEISS), Simmons).

While some of the categorization schemes above are intended for exposure, little study has been done to evaluate how varying categorization may affect the predictive power of exposure models. The more detailed and accurate mechanistic models of chemical exposure (e.g. EPA's Stochastic Human Exposure and Dose Simulation or SHEDS) require parameters that are variable within a single category in the exposure focused schemes noted above. Finding an appropriate categorization scheme that provides the flexibility for modeling at multiple resolutions will take time and study. Until then, we advocate for documenting products electronically with more detailed categorization schemes rather than less detailed schemes as aggregating data for a future classification paradigm is facile (requiring only mapping between the detailed categories and the new system) while partitioning is taxing (requiring manual assessment for each product as to which more specialized class it belongs).

The categories contained in the retailer's marketing system are by necessity fairly detailed, allowing consumers to navigate the tree to their products of interest. This makes it suitable as an initial classification for use in CPCPdb. However, while the utility of this classification in retail is obvious, some are incorrect for exposure prioritization. Take for example the product "Andis Cool Care Plus", a spray used to clean electric hair clippers, which is categorized under "Beauty; Hair Styling Tools". While this category is appropriate to the retailer and the consumers, an exposure scientist may consider changing the category to "cleaner".

Another more generic difficulty for product categorization can be viewed when looking at "Fizzies Fish 'n Splash Asst Holiday 2007". This package contains a toy (plastic fishing rod), bathing accessories (bath fizzies), and a body cleanser (moldable foam soap). The proper categorization for a combination product (or

multiple component kit) of this type is difficult to discern either programmatically or during typical curation efforts. These examples highlight the need to generate a standard for categorization of consumer products in the realm of exposure sciences. Determining the appropriate way to deal with the complexity of consumer product classifications is something upon which we will devote additional consideration in the future.

5. Conclusion

More than 80% (65 of 83 substances) of the chemicals listed in the EPA's Toxic Substances Control Act (TSCA) priority workplan are known to be present in consumer products http://www.epa.gov/oppt/existingchemicals/pubs/Work_Plan_Chemicals_Web_Final.pdf. Knowledge of consumer product chemical composition (both type and quantity) or product formulations is vital to ensuring public safety as it fills in an important data gaps in near field exposure models. By collecting consumer product ingredient information and product use category information in a centralized repository, we have a key source that can inform multiple research initiatives intent upon better understanding of near field chemical exposures. In parallel with the toxicoinformatic infrastructures required for computational toxicology, these data bring informatics-driven exposure prioritization one step closer to high-throughput risk assessment.

Conflict of Interest

The authors declare there are no conflicts of interest.

Transparency document associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fct.2013.12.029.

Disclaimer

The United States Environmental Protection Agency has provided administrative review and has approved this manuscript for publication. The views expressed in this manuscript are those of the authors and do not necessarily reflect the views or policies of the Agency.

Acknowledgements

The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. The initial funding for this research came from a U.S. EPA Pathfinder Innovation Project award for "Systems Reality Modeling." We are grateful to the Shaw University Research Internship program for providing students that assisted in this research. We thank Charles Bevington and Cathy Fehrenbacher of the U.S. EPA's Office of Chemical Safety and Pollution Prevention for collaboration and helpful discussions. We thank Mike Uhl (Lockheed Martin Information Technology Services), Ravi Nair and Heidi Paulson (both U.S. EPA's Environmental Modeling and Visualization Laboratory) for project coordination on the interface design. We thank Pertti Hakkinen (National Library of Medicine) and Henry Delima (Henry Delima Associates) for providing input through ongoing discussion. We thank Linda Sheldon, Satori Marchitti and Haluk Ozkaynak (NERL) for administrative review and feedback.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fct.2013.12.029.

References

Al-Saleh, I., Shinwari, N., Alsabbaheen, A., 2011. Phthalates residues in plastic bottled waters. J. Toxicol. Sci. 36 (4), 469–478.

Anastas, P., Teichman, K., Hubal, E.C., 2010. Ensuring the safety of chemicals. J. Eposure Sci. Environ. Epidemiol. 20 (5), 395–396.

Berellini, G., Springer, C., Waters, N.J., Lombardo, F., 2009. In silico prediction of volume of distribution in human using linear and nonlinear models on a 669 compound data set. J. Med. Chem. 52 (14), 4488–4495.

Bracken, M.C., Weiss, I.J., 1977. Database development in a regulatory agency. J. Chem. Inf. Comput. Sci. 17 (4), 202–205.

Byer, W.L., Landau, H.B., Neufeld, M.L., Rosenthal, H., 1976. Building a chemical ingredient data base for industrial and consumer products. J. Chem. Inf. Comput. Sci. 16 (3), 137–141.

Dodson, R.E., Nishioka, M., Standley, L.J., Perovich, L.J., Brody, J.G., Rudel, R.A., 2012. Endocrine disruptors and asthma-associated chemicals in consumer products. Environ. Health Perspect. 120 (7), 935.

Egeghy, P.P., Vallero, D.A., Cohen Hubal, E.A., 2011. Exposure-based prioritization of chemicals for risk assessment. Environ. Sci. Policy 14 (8), 950–964.

Glegg, G.A., Richards, J.P., 2007. Chemicals in household products: problems with solutions. Environ. Manage. 40 (6), 889–901.

Goldsmith et al., 2012. Summary Report for "Personal Chemical Exposure Informatics: Visualization and Exploratory Research in Simulations and Systems (PerCEIVERS)", EPA report EPA/600/R13/041.

Gosselin, R.E., Smith, R.P., Hodge, H.C., 1984. Clinical Toxicology of Commercial Products. Williams & Wilkins.

Hertwich, E.G., 2005. Life cycle approaches to sustainable consumption: a critical review. Environ. Sci. Technol. 39 (13), 4673–4684.

Jayjock, M.A., Chaisson, C.F., Franklin, C.A., Arnold, S., Price, P.S., 2008. Using publicly available information to create exposure and risk-based ranking of chemicals used in the workplace and consumer products. J. Eposure Sci. Environ. Epidemiol. 19 (5), 515–524.

Judson, R.S., Martin, M.T., Egeghy, P., Gangwal, S., Reif, D.M., Kothiya, P., Richard, A.M., 2012. Aggregating data for computational toxicology applications: the US environmental protection agency (ePA) Aggregated Computational toxicology Resource (ACtoR) system. Int. J. Mol. Sci. 13 (2), 1805–1831.

Kelly, B.C., Ikonomou, M.G., Blair, J.D., Morin, A.E., Gobas, F.A., 2007. Food web-specific biomagnification of persistent organic pollutants. Science 317 (5835), 236–239.

Klepeis, N.E., Nelson, W.C., Ott, W.R., Robinson, J.P., Tsang, A.M., Switzer, P., Engelmann, W.H., 2001. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. J. Expo. Anal. Environ. Epidemiol. 11 (3), 231–252.

Lai, A.C., Thatcher, T.L., Nazaroff, W.W., 2000. Inhalation transfer factors for air pollution health risk assessment. J. Air Waste Manage. Assoc. 50 (9), 1688–1699.

Little, J.C., Weschler, C.J., Nazaroff, W.W., Liu, Z., Cohen Hubal, E.A., 2012. Rapid methods to estimate potential exposure to semivolatile organic compounds in the indoor environment. Environ. Sci. Technol. 46 (20), 11171–11178.

Meyer, E.C., Slick, S.K., 2001. Consumer and Industrial goods. In: Kaliski, Burton S. (Ed.), Encyclopedia of Business and Finance. New York: Macmillan Reference USA.

Morgan, M.K., Sheldon, L.S., Croghan, C., Chuang, J.C., Lyu, C., Wilson, N.K., Lordo, R.A., Brinkman, M.C., Morse, N., Chou, Y.L., Hamiltin, C., Finegold, J.K., Hand, K., Gordon, S.M., 2005. A PILOT STUDY OF CHILDREN'S TOTAL EXPOSURE TO PERSISTENT PESTICIDES AND OTHER PERSISTENT ORGANIC POLLUTANTS (CTEPP). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-04/193 (NTIS PB2006-101286).

Nazaroff, W., Weschler, C.J., Little, J.C., Hubal, E.A.C., 2012. Intake to production ratio: a measure of exposure intimacy for manufactured chemicals. Environ. Health Perspect. 120 (12), 1678.

National Research Council, 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press. <http://www.nap.edu/catalog.php?record_id=11970>.

National Research Council, 2012. Exposure Science in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press. <http://www.nap.edu/catalog.php?record_id=13507>.

NTP, 2002. National Toxicology Program: Annual Plan for Fiscal Year 2001. Research Triangle Park, NC: National Toxicology Program.

Organisation for Economic Co-operation and Development (OECD), 2012. Crosswalk of Harmonized U.S. – Canada Industrial Function and Consumer and Commercial Product Categories with EU Chemical Product and Article Categories. Environment Directorate, Series on Testing and Assessment No. 167. ENV/JM/MONO(2012)5.

Pellizzari, E.D., Perritt, R.L., Clayton, C.A., 1999. National human exposure assessment survey (NHEXAS): exploratory survey of exposure among population subgroups in EPA Region V. J. Expo. Anal. Environ. Epidemiol. 9 (1), 49.

Rudel, R.A., Camann, D.E., Spengler, J.D., Korn, L.R., Brody, J.G., 2003. Phthalates, alkylphenols, pesticides, polybrominated diphenyl ethers, and other endocrine-disrupting compounds in indoor air and dust. Environ. Sci. Technol. 37 (20), 4543–4553.

Sanderson, H., Greggs, W., Cowan-Ellsberry, C., DeLeo, P., Sedlak, R., 2013. Collection and dissemination of exposure data throughout the chemical value chain: a case study from a global consumer product industry. Human Ecol. Risk Assess.: Int. J. 19 (4), 999–1013.

Sarver, J.G., White, D., Erhardt, P., Bachmann, K., 1997. Estimating xenobiotic half-lives in humans from rat data: influence of log P. Environ. Health Perspect. 105 (11), 1204.

Schettler, T., 2006. Human exposure to phthalates via consumer products. Int. J. Androl. 29 (1), 134–139.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome. research. 13 (11), 2498–2504.

Shin, H.M., McKone, T.E., Bennett, D.H., 2012. Intake fraction for the indoor environment: a tool for prioritizing indoor chemical sources. Environ. Sci. Technol. 46 (18), 10063–10072.

Shin, H.M., McKone, T.E., Nishioka, M.G., Fallin, M.D., Croen, L.A., Hertz-Picciotto, I., Bennett, D.H., 2013. Determining Source Strength of Semivolatile Organic Compounds using Measured Concentrations in Indoor Dust. Indoor air.

Smith, R., 2007, September. An Overview of the Tesseract OCR Engine. In: Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on IEEE, vol. 2, pp. 629–633.

Vallero, D., 2010. Environmental Biotechnology: A Biosystems Approach. Academic Press.

Wallace, L.A., 1987. The total exposure assessment methodology (TEAM) study: summary and analysis: volume I. Off. Res. Develop. US Environ. Protect. Agency 600, 6–87.

Wambaugh, J.F., Setzer, R.W., Reif, D.M., Gangwal, S., Mitchell-Blackwood, J., Arnot, J.A., Joliet, O., Frame, A., Rabinowitz, J.R., Knudsen, T.B., Judson, R.S., Egeghy, P., Vallero, D.A., Cohen Hubal, E.A., 2013. High Throughput Models for Exposure-Based Chemical Prioritization in the ExpoCast Project. Environ. Sci. Technol..

Wenger, Y., Li, D., Jolliet, O., 2012. Indoor intake fraction considering surface sorption of air organic compounds for life cycle assessment. Int. J. Life Cycle Assess. 17 (7), 919–931.

Weschler, C.J., Nazaroff, W.W., 2008. Semivolatile organic compounds in indoor environments. Atmos. Environ. 42 (40), 9018–9040.

Weschler, C.J., 2009. Changes in indoor pollutants since the 1950s. Atmos. Environ. 43 (1), 153–169.

Weschler, C.J., Nazaroff, W.W., 2010. SVOC partitioning between the gas phase and settled dust indoors. Atmos. Environ. 44 (30), 3609–3620.

Wilke, O., Jann, O., Brödner, D., 2004. VOC-and SVOC-emissions from adhesives, floor coverings and complete floor structures. Indoor Air 14 (s8), 98–107.